Journal of Medical Virology 73:123–130 (2004)

PCR-Probe Capture Hybridization Assay and Statistical Model for SEN Virus Prevalence Estimation

Yasuhito Tanaka,[1,3] Ruth Pfeiffer,[2]* Anthony E.T. Yeo,[1,4] Masashi Mizokami,[3] Brian R. Edlin,[5,6] Philip Castle,[2] Thomas R. O'Brien,[2] Harvey J. Alter,[1] and J. Wai-Kuo Shih[1]

[1]*Department of Transfusion Medicine, Warren Grant Magnuson Clinical Center, The National Institutes of Health, DHHS, Bethesda, Maryland*
[2]*Division of Cancer Epidemiology and Genetics, National Cancer Institute, The National Institutes of Health, DHHS, Bethesda, Maryland*
[3]*Department of Clinical Molecular Informative Medicine, Nagoya City University Graduate School of Medical Sciences, Nagoya, Japan*
[4]*Jacobus Pharmaceutical Company, Inc., Princeton, New Jersey*
[5]*Institute for Health Policy Studies, University of California, San Francisco, California*
[6]*Center for the Study of Hepatitis C, Weill Medical College of Cornell University, New York, New York*

SEN viruses (SENV) are newly discovered blood-borne single-stranded circular DNA viruses that may play a role in liver disease. To date, no serologic assays are available for the detection of SENV antigens or antibodies. We report on a rapid and sensitive molecular assay for the detection of four SENV strains (SENV-A, -C, -D, -H). This method uses PCR with universal primers and microwell capture hybridization with type-specific probes. Cut-off points to define ''infected'' based on chemiluminescence readings were determined from a statistical mixture model applied to samples from 300 injection drug users (IDUs) in San Francisco. Based on the estimated cut-off points, we examined the prevalence of SENV infection among 232 healthy US blood donors and assessed sensitivity and specificity of the assay in a small validation sample of infected individuals with partial sequence information. ***J. Med. Virol. 73:123–130, 2004.***

KEY WORDS: injection drug users; cut-off value; sensitivity; SENV; mixture model

INTRODUCTION

A novel single-stranded DNA virus of about 3,800 nucleotides was discovered in the serum of an HIV-1-infected injection drug user (IDU) from Italy in 2000 [Primi et al., 2000; Sottini et al., 2001; Tanaka et al., 2001]. This agent was designated SEN virus (SENV) based on the initials of this patient. There is evidence of at least seven SENV genotypes, SENV-A, B, C, D, E, F, G, and H [Primi et al., 2000; Sottini et al., 2001]. The genome structure of SENV is related to the genome structure of the chicken anemia virus in the genus *Gyrovirus* as well as to the TT virus (TTV), however, the sequence similarity is minimal except for a short stretch at 3,816–3,851 of TA278. Currently the International Committee on Taxonomy of Viruses (ICTV) working group is proposing the full name for the TTV as TorqueTenoVirus in a new genus *Anellovirus* (ring) [Hino, 2002].

SENV strains appear to be common in people who are exposed to exogenous blood, but infrequent in those who are not [Umemura et al., 2001; Wilson et al., 2001]. Infection with SENV can be chronic [Umemura et al., 2001], and re-infection with SENV strains can occur [Wilson et al., 2001]. The clinical importance of SENV infection is uncertain. There is evidence suggesting that SENV causes acute hepatitis in some people who become infected with this virus [Shibata et al., 2001]. One study found that persons infected with SENV are less likely to respond to treatment for HCV [Rigas et al., 2001], but that finding has not been confirmed.

To date there are no serologic tests to detect SENV antigen or anti-viral antibody. We developed a group of PCR-probe capture hybridization assays to detect DNA from strains of SENV. Our assays are a rapid and robust

Yasuhito Tanaka and Ruth Pfeiffer contributed equally to this paper.

*Correspondence to: Ruth Pfeiffer, Division of Cancer Epidemiology and Genetics, NCI, The National Institutes of Health, 6120 Executive Blvd., Room 8030, Rockville, MD 20852.
E-mail: pfeiffer@mail.nih.gov

Accepted 18 December 2003
DOI 10.1002/jmv.20067
Published online in Wiley InterScience (www.interscience.wiley.com)

alternative to cloning and sequencing viral genes. The tests rely on the hybridization of biotinylated SENV probes to amplified SENV targets labeled with digoxigenin, capture of hybrids by strepavidin-coated 96-well ELISA plates, and detection of hybrids using alkaline phosphatase conjugated anti-digoxigenin antibodies and chemiluminescent substrate. Because there is no gold standard to define who is "SENV-infected," we used a statistical mixture model to determine cut-off points for positive results. This model assumes that two different populations, one infected and one uninfected, give rise to a bimodal distribution of assay readings. The optimal cut-off points maximize sensitivity and specificity as determined from the mixture model fitted to samples from IDUs. We assessed the sensitivity and specificity of the assays by applying the estimated cut-off points to assay measurements obtained from a small validation sample with partial sequencing information. We also extended this model to incorporate auxiliary information, such as sequence homology data, to improve estimation of the cut-off points. Using this approach, we investigated the prevalence of SENV infection in a group of blood donors.

MATERIALS AND METHODS

Subjects and Specimens

Serum samples were obtained from 300 IDUs in San Francisco, CA [Atkinson et al., 2003] and 232 US blood donors at the National Institutes of Health (NIH), Bethesda, MD, who had been screened for other parenteral exposures (i.e., HIV, HBV, and HCV). All study subjects gave written informed consent. All serum samples were stored at −70°C.

A validation sample of 45 specimens that had been sequenced was available: 21 samples (obtained from IDUs) that contained the sequence of only one SENV strain, and 24 cloned PCR fragments (obtained from eight IDUs and 16 liver disease patients). We also used data on 150 negative samples: 66 PCR-negative samples obtained from blood donors, 20 commercial negative controls (Diasorin, Inc., Saluggia, Italy), 30 negative controls for PCR (reagents only), 18 hybridization assay-negative controls (sterile water, substrate blank), and 16 cloned PCR fragments of different type from each type-specific probe.

Detection of SEN-V DNA

Nucleic acids were extracted from 100 μl of serum by QIAamp DNA Blood Mini Kit (Qiagen, Valencia, CA). The extracted DNA was eluted in 160 μl buffer containing 10 mM Tris-HCl, 0.5 mM EDTA, pH 9.0. The presence of SENV DNA was detected by polymerase chain reaction (PCR). As previously reported [Primi et al., 2000], universal primers for only SENV designated NEW BCD 1S (SEQ ID NO: 115) and L 2AS (SEQ ID NO: 71), derived from conserved sequences within partial open reading frame 1 among the SENV strains, SENV-A, SENV-B, SENV-C, SENV-D, SENV-F, SENV-G, and SENV-H. In order to improve sensitivity, primers NEW BCD 1S and L 2AS were slightly modified on the basis of additional sequence data and the resulting primer NEW BCD 1S′ (sense primer 5′-CCCAAACTRTTTGAACMASTGGTA-3′, R = A or G, M = A or C, S = C or G) was used in combination with NEW L 2AS (anti-sense primer 5′-CCTCGGTTKSAAAKGTYTGATAGTG-3′, K = G or T, Y = C or T). PCR reactions were performed as a 40 μl PCR mixture containing Gold PCR buffer with $MgCl_2$, 0.4 μM of each PCR primer, digoxigenin labeling mix (200 μmol/l of dATP, dCTP, dGTP each, 190 μmol/l dTTP and 10 μmol/l digoxigenin-11-dUTP), 1.25 U of AmpliTaq Gold DNA polymerase (Applied Biosystems, Foster City, CA) and 10 μl of extracted DNA. The reaction was performed in GeneAmp® PCR System 9700 (Applied Biosystems) without mineral oil. The PCR program consisted of pre-heating at 94°C for 10 min, 50 cycles of 94°C for 1 min, 55°C for 1 min, and 72°C for 1 min, and incubation at 72°C for 9 min. PCR products (SENV-A, 208 bp, SENV-D, 229 bp, SENV-C or -H, 226 bp excluding primers sequences) were analyzed by chemiluminescence (see Materials and Methods below) with each SENV type-specific 5′-end biotinylated probe that targeted conserved sequences within the amplified DNA (Fig. 1): SENV-A; 5′-CCCCATGAAAGGGGAAGAGGCCTACACTGACTTT-3′, SENV-D; 5′-ATGATAGGCTTCCCYTTTAACTATAACCCA-3′ (Y = C or T), SENV-H; 5′-CCAGTAATAGGCACTTCTGCTTTAGAACA-3′, SENV-C/H; 5′-CCCCTTCCAGGTATTGCATGAAGAGTATTAC-3′.

SENV probes were based on the Diasorin patent by Primi et al. [2000], with SENV-A and SENV-H probes slightly modified from the patent to improve type specificity. Previous phylogenetic analysis [Tanaka et al., 2001] showed SENV-C and SENV-H belonged to the same genotype with 78.3% nucleotide homology within total ORF1. Thus, data on SENV-C and -H were combined and the term of SENV-C/H denotes infection with either SENV-C or -H.

Streptavidin-coated black ELISA plates (Roche Molecular Biochemicals, Indianapolis, IN) were incubated overnight at 4°C with each biotinylated probe (10 ng/well). After washing the plate, a 100 μl of 90 μl of hybridization solution and 10 μl of PCR-amplified products, denatured at 95°C for 5 min, was added to each well. Hybridization was performed at 50°C for 1 h for SENV-A and SENV-D probes and at 45°C for 1 h for SENV-H and SENV-C/H probes. ELISA plates were then washed, and anti-digoxigenin F(ab) fragments conjugated to alkaline phosphatase (Boehringer Mannheim, Mannheim, Germany) were added to the wells (100 μl/well). The mixture was then incubated for 30 min at 37°C. Finally, after washing, 0.25 mM disodium 3-(4-methoxyspiro {1, 2-dioxetane-3, 2′-(5′-chloro) tricyclo [3.3.1.1^{3,7}] decan}-4-yl) phenyl phosphate (CSPD®, Boehringer Mannheim) substrate in 0.1 M Tris-HCl, 0.1 M NaCl, pH 9.5 was added (100 μl/well) to the wells for chemiluminescence signal development. Chemiluminescence was read in a 1450 Micro Beta PLUS scintillation counter (Wallac, Inc., Turku, Finland) and raw data were shown as counts per second.

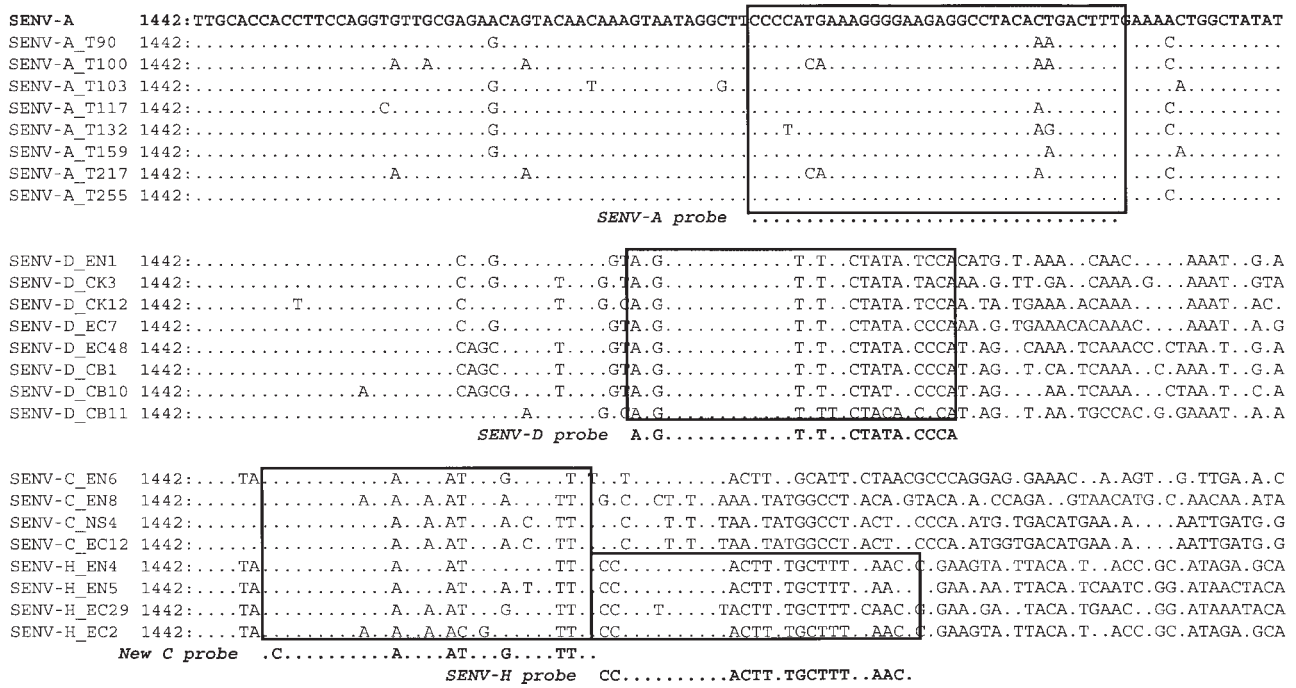

Fig. 1. Alignment of SEN viruses (SENV) clones and location hybridized with each probe. These sequences were aligned on prototype SENV-A and the number indicates the nucleotide position of SENV-A. Each box is the location hybridized with each probe.

Cloning and Sequencing of PCR Amplified SENV DNA Products

Sequences of each isolate were determined directly or after cloning into the pCR2.1-TOPO vector (Invitrogen, San Diego, CA). The sequencing reactions were performed with Big Dye (Perkin-Elmer Applied Biosytems, Foster City, CA) in GeneAmp PCR System 9700 and analyzed in ABI 310 DNA automated sequencer according to the manufacturer's protocol.

Determining Cut-Off Points for the Assay

Model 1. We relied on two-component statistical mixture models to determine optimal cut-off points for the chemiluminescence readings obtained for each SENV developed by [Pfeiffer et al., 1999]. These models assume that the measurements arise from two different populations, an infected and an uninfected one, that yield a bimodal distribution of signal intensity. The measurements of each population are modeled by a different probability density, i.e., the derivative of the probability distribution function, each with its own set of parameters. A third model parameter is the mixing proportion, which is the proportion of truly infected individuals in the sample. Details on the model and parameter estimation are given in Appendix.

We used the IDU samples to estimate the cut-off points for each strain of SENV. First, we examined histogram plots to determine whether a statistical transformation was needed to create a probability density mixture consisting of two normal protbability densities. The normalizing transformation was $\log_e(x+0.5)$, where x denoted the chemiluminescence reading. If an individual had repeated readings, x was the mean of all readings. The optimal cut-off point maximized the sum of sensitivity and specificity as determined from the fitted mixture model. We obtained 95% confidence intervals (95% CI) for the cut-off points and all prevalence estimates using a bootstrap procedure with 1,000 repetitions.

Using the cut-off points from the IDU data, we estimated prevalence of SENV infection in a sample of US blood donors, who we presume to be at low risk of infection because they are screened for other blood borne viruses. We also estimated sensitivity and specificity of the assays for the small validation sample of 45 specimens that had been sequenced. We defined a sample as positive if a strain-specific nucleotide sequence was present.

Model 2. Model 2 is a mixture model designed to improve the estimation of the cut-off points by using a logistic regression function to incorporate homology data into the modeling of the mixing probabilities. Details are given in the Appendix. Model 2 was applied to the validation data. The classification of the samples based on Model 2 is listed in Table I.

Model 3. We also estimated the cut-off points by calculating the mean + 2 SD (standard deviations) of the log-transformed data on the 150 negative samples. The mean + 2 SD approach has been used previously to estimate cut-off points for other assay measurements

TABLE I. Percent Nucleotide Homology Among SEN Viruses (SENV) Strains and Four Probes and Results on the Basis of Cut-Off Points by Statistical Models

Specimen	Strain with highest % homology[a]	Homology with probes			
		SENVA (34)[c]	SENVD (30)[c]	SENVH (29)[c]	SENV-C/H (31)[c]
Injection drug user (IDU) clone SENV-A_T90	A:97.1%	94%	57%	59%	74%
IDU clone SENV-A_T100	A:95.2%	88%	53%	52%	74%
IDU clone SENV-A_T103	A:97.1%	100%	53%	45%	71%
IDU clone SENV-A_T117	A:98.1%	100%	63%	59%	71%
IDU clone SENV-A_T132	A:97.6%	91%	53%	52%	74%
IDU clone SENV-A_T159	A:96.2%	88%	63%	59%	77%
IDU clone SENV-A_T217	A:96.2%	91%	53%	52%	77%
IDU clone SENV-A_T255	A:98.6%	100%	57%	59%	77%
IDU 3	A:94.2%	97%	53%	59%	77%
IDU 11	A:97.2%	97%	53%	59%	77%
IDU 13	A:96.2%	100%	53%	59%	74%
IDU 15	A:92.7%	91%	47%	59%	77%
IDU 16	A:95.5%	97%	53%	59%	77%
IDU 20	A:95.8%	97%	57%	34%	74%
IDU 21	A:95.0%	94%	57%	34%	74%
IDU 22	A:95.6%	91%	53%	34%	74%
IDU 28	A:95.8%	91%	53%	34%	74%
IDU 268	A:96.9%	97%	53%	59%	74%
IDU 278	A:96.0%	97%	53%	59%	77%
Clone SENV-D_EN1[b]	D:88.6%	35%	97%	52%	74%
Clone SENV-D_CK3[b]	D:87.8%	35%	93%	52%	77%
Clone SENV-D_CK12[b]	D:89.5%	35%	97%	52%	77%
Clone SENV-D_EC7[b]	D:86.9%	53%	100%	52%	74%
Clone SENV-D_EC48[b]	D:87.3%	47%	100%	52%	71%
Clone SENV-D_CB1[b]	D:88.6%	47%	100%	52%	71%
Clone SENV-D_CB10[b]	D:86.9%	47%	97%	52%	71%
Clone SENV-D_CB11[b]	D:86.0%	38%	90%	55%	74%
IDU 272	D:86.2%	47%	93%	55%	68%
IDU 451	D:83.0%	38%	83%	48%	74%
Clone SENV-C_EN6[b]	C:81.5%	35%	73%	69%	90%
Clone SENV-C_EN8[b]	C:76.6%	38%	43%	45%	87%
Clone SENV-C_NS4[b]	C:76.7%	41%	43%	45%	87%
Clone SENV-C_EC12[b]	C:76.1%	41%	43%	45%	87%
IDU 10	C:90.3%	50%	50%	66%	90%
IDU 24	C:73.3%	35%	43%	38%	87%
IDU 83	C:79.4%	47%	47%	48%	94%
Clone SENV-H_EN4[b]	H:92.5%	53%	50%	100%	94%
Clone SENV-H_EN5[b]	H:91.3%	53%	50%	97%	90%
Clone SENV-H_EC29[b]	H:89.4%	56%	50%	90%	94%
Clone SENV-H_EC2[b]	H:90.7%	53%	50%	100%	81%
IDU 8	H:90.2%	32%	67%	97%	90%
IDU 85	H:85.1%	47%	50%	93%	90%
IDU 193	H:91.8%	35%	50%	93%	97%
IDU 214	H:86.3%	47%	50%	97%	97%
IDU 233	H:91.4%	32%	47%	97%	97%

[a]The highest percent homology among SENV strains and each prototype SENV (SENV-A, -C, -D, and -H).
[b]Clones from liver disease patient.
[c]Number of oligonucleotides of each probe.

[Meden et al., 1994; Zhang et al., 2000]. The probability that a normally distributed quantity exceeds the mean + 2 SD is roughly 2.3% [Armitage and Berry, 1994]. Taking the cut-off point to be the mean + 2 SD thus reflects the understanding that the largest 2.3% of the assay measurements will be misclassified as infected, in other words, the maximum specificity of the assay is set at 97.7%.

All analyses were conducted using MATLAB 5.0 [The Mathworks, Inc., 1999], as described in Pfeiffer et al. [1999].

RESULTS

Development of SENV PCR-Microwell Hybridization Assay

For SENV-A and SENV-D, the high homology corresponded to high chemiluminescence readings (positive), which were clearly distinguished from lower readings corresponding to lower homology. A previous phylogenetic analysis [Tanaka et al., 2001] indicated that the homology between SENV-C and -H was 78.3%, and that these strains were significantly more

homologous with each other than with any other pairs. Thus, the SENV-C/H probe was not able to distinguish between SENV-C and -H. As SENV-C and -H strains had greater genetic diversity within a genotype than that within the other genotypes, we combined results for the SENV-H probe and the SENV-C/H probe to decrease the false negativity of the detection. Of 16 SENV-C or -H specimens, seven SENV-C specimens had low homologies with the SENV-H probe and one SENV-H specimen (EC2) had relatively low homology (81%) with the SENV-C/H probe (Table I). This low homology resulted in lower chemiluminescence readings; several readings appeared to be borderline making it difficult to define infected and uninfected subjects. Thus, there are likely to be a few unclassifiable data; however, most results were also consonant with the sequence data (Table I). These two different probes were useful in detecting the divergent strains in SENV-C/H.

Cut-Off Values and Prevalence Estimates Among IDUs and Blood Donors

Figure 2 shows the histogram of the SENV-A data, the fitted mixture model, and the cut-off point on the log transformed scale, $1.65 = \log(4.69 + 0.5)$. The individuals with measurements to the right of the cut-off point are classified as "infected," the others are classified as "uninfected." Similar histograms were observed for the other SENV strains. Table II presents cut-off points and prevalence estimates along with their 95% CIs. The estimated prevalence of SENV-A was 52.3% (95% CI: 49.3–61.0%). For SENV-D the estimated prevalence was 11.0% (95% CI 7.0–14.0%) and for SENV-H 34.3% (95% CI 25.3–36.0%). For SENV-C/H, the prevalence estimate was 51.5% (95% CI: 45.5–57.2%).

We used the cut-off points obtained from the IDU data to estimate the prevalence of SENV strains among 232 blood donors (Table II). The estimated prevalence rates were: SENV-A, 1.3% (95% CI: 0–3.0%), SENV-D, 1.7% (95% CI: 0.4–3.5%), SENV-H 9.1% (95% CI: 5.6–13.4%), and SENV-C/H 7.8% (4.3–11.2%).

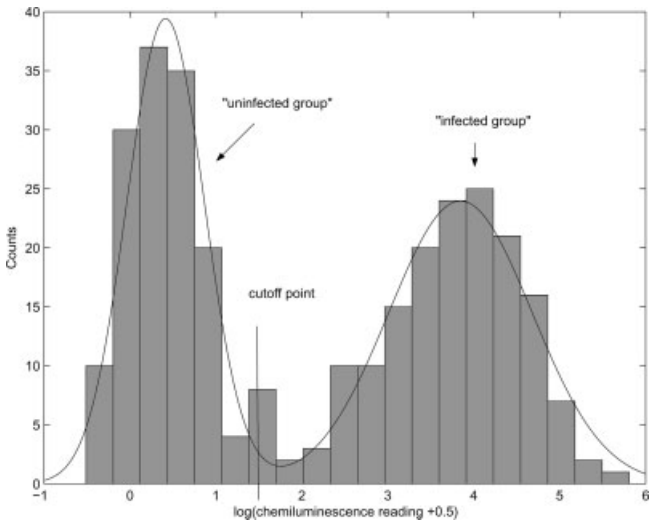

Fig. 2. SENV-A measurements among injection drug users (IDUs). Smooth line indicates the fitted mixture model.

We also fitted Model 1 to the US blood donor samples to obtain "internal cut-off points" for that population. For SENV-A, this model resulted in a cut-off value of 3.5 and a slightly higher prevalence estimate (2.6%), which was not significantly different from the estimate obtained from applying the cut-off points from the IDU sample (1.3%). The internal cut-off points of 6.4, 7.4, and 8.4 for SENV-D, -H, and -C/H, respectively (data not shown), differed slightly from the cut-off points based on the IDU data, but these differences were not statistically significant, and the prevalence estimates did not change. The prevalence estimates were not sensitive to these changes in cut-off points because the assay readings for the infected and uninfected populations were well separated.

Validation Analysis

The 45 samples with a single SENV strain were used to validate sensitivity and specificity of the assays for various choices of cut-off points. The true status of infection was defined by presence of a virus-specific nucleotide sequence. The classification of the individual samples is given in Table I.

Model 1, which assumes constant mixing probabilities, yielded these cut-off points (95% CI) given in Table III: SENV-A, 9.5 (6.4–14.1); SENV-D, 5.6 (3.7–15.2); SENV-H, 6.1 (4.2–13.2), and SENV-C/H, 17.1 (6.7–33.3). The wide confidence intervals reflect the small sample size, with the cut-off point for the SENV-C/H probe being the least precise. The differences between the Model 1 cut-off points from the validation data and the cut-off points from the IDU data are not statistically significant.

Model 2, an extension of the mixture model that incorporated homology information of all possible SENV strains, resulted in identical cut-off points for SENV-A and SENV-D, but slightly different cut-off points for SENV-H probe (6.6) and the SENV-C/H probe (15.1) (Table III). The differences between the Model 1 cut-off points from the validation data and the cut-off points from the IDU data are not statistically significant.

The cut-off points from Model 3, obtained by calculating the mean + 2 SD of 150 negative samples, were close to the IDU cut-offs for SENV-D, and close to the cut-offs from Model 1 for SENV-C/H. For SENV-A, the cut-off value from Model 3 was in between the IDU cut-off (4.8) and the Model 1 and Model 2 cut-off (9.5). For SENV-H, Model 3 resulted in a much higher cut-off value (11.4) than the ones obtained from either Model 1 or 2 (6.1 and 6.6, respectively), and the IDU cut-off value (4.6).

Table III also shows the estimates of sensitivity and specificity based on the validation sample with 95% CIs for all choices of cut-off values. All assays are highly sensitive for all SENV strains using any of the cut points (92–100%) and the assays for SENV-A, -D, and -H was also highly specificity (92–100%). The assay for SENV-C/H was highly specific (100%) with the higher cut

TABLE II. Cut-Off Points Obtained From Model 1 (Applied to IDU Data) and Resulting Estimated Prevalence of SENV Strains Among Blood Donors and IDUs (95% CI in Parenthesis)

Strain	Cut-off	Prevalence for IDU data (n = 300)	Prevalence for blood donor data (n = 232)
SENV-A	4.8 (3.7–6.1)	52.3% (49.3–61.0%)	1.3% (0–3.0%)
SENV-D	4.5 (2.7–4.4)	11.0% (7.0–14.0%)	1.7% (0.4–3.5%)
SENV-H	4.6 (3.4–6.0)	34.3% (25.3–36.0%)	9.1% (5.6–13.4%)
SENV-C/H	5.0 (1.6–5.4)	51.5% (46.2–57.5%)	7.8% (4.3–11.2%)

points established using the validation samples but was not specific (62%, 95% CI 44–80%) when using the lower cut point established for the IDU samples.

DISCUSSION

PCR-capture hybridization assays have been used to detect specific strains of several viral families, for example, human papillomavirus (HPV) types 16, 18, and 33 (3). This typing approach is more sensitive than agarose gel electrophoretic analysis in detecting amplified PCR products [Andreoletti et al., 1996], and more convenient than sequencing analyses. We developed a PCR amplification followed by a microwell capture hybridization assay using chemiluminescent detection for SENV-A, SENV-D, and SENV-C/H. The method described in this paper is also more convenient and efficient than the initial commercial detection system DEIA^{TM} (Diasorin, Inc.), which consists of a PCR assay with immunoassay (data not shown). The assay threshold for detection of the PCR-microwell hybridization is in the order of 10^2–$10^{2.5}$ copies/ml.

To determine cut-off points for the assays, statistical mixture models were fitted to a sample of IDU data, a population in which prevalences were expected to be high. The optimal cut-off points were the values that maximized the sum of sensitivity and specificity. Model 1 assumed constant mixing probabilities while Model 2 allowed the incorporation of homology information to improve the estimation of the cut-off points. The cut-off points estimated from the Models 1 and 2 were close to cut-off points that were obtained from Model 3, by calculating the mean + 2 SD of 150 negative samples. A limitation is that Model 3 assumes a specificity of 97.7%, whereas the mixture models determine cut-off points that maximize the sum of sensitivity and specificity without assuming fixed values.

When applied to a sample of 232 US blood donors, who were screened for other parenterally transmitted viruses (HCV, HBV, and HIV), the prevalence estimates for SENV-A (1.3%) and SENV-D (1.7%) were low, as expected. The prevalence of SENV-C/H among the blood donors (7.8%) was significantly higher than the prevalence of SENV-A or -D. This difference may be attributed to the lower specificity of the C/H probe, which was estimated to be around 60%.

As noted above, all positive controls of each test gave high values, and all negative controls and clones

TABLE III. Cut-Off Points, Sensitivity, and Specificity for the 45 Validation Samples With 95% CIs in Parenthesis

	Strain (number infected)			
	SENV-A (19/45)	SENV-D (10/45)	SENV-H (9/45)	SENV-C/H (7/45)
Model 1				
Cut-off	9.5 (6.4–14.1)	5.6 (3.7–15.2)	6.1 (4.2–13.2)	17.1 (6.7–33.3)
Sensitivity	100% (85–100%)	100% (74–100%)	100% (71–100%)	94% (82–100%)
Specificity	100% (89–100%)	100% (91–100%)	92% (83–96%)	100% (92–100%)
Model 2				
Cut-off	9.5 (6.4–14.1)	5.6 (3.7–15.2)	6.6 (4.8–14.2)	15.1 (5.4–30.2)
Sensitivity	100% (89–100%)	100% (91–100%)	92% (83–96%)	100% (92–100%)
Specificity	100% (89–100%)	100% (91–100%)	92% (83–96%)	100% (92–100%)
Model 3				
Cut-off	7.6	4.6	11.4	15.5
Sensitivity	100% (89–100%)	100% (91–100%)	92% (83–96%)	100% (92–100%)
Specificity	100% (89–100%)	100% (91–100%)	94% (86–100%)	100% (92–100%)
IDU cut-off points				
Cut-off	4.8	4.5	4.6	5.0
Sensitivity	100% (89–100%)	100% (91–100%)	100% (80–100%)	100% (92–100%)
Specificity	96% (88–100%)	100% (91–100%)	94 % (86–100%)	62% (44–80%)

Model 1, assumes a cut-off constant probability of infection.
Model 2, incorporates homology information into the probability of infection.
Model 3, defines the cut-off points to be the mean + 2 SD for 150 negative samples.

produced low values, although the statistical models resulted in a few false positives classifications. There were also some misclassifications for SENV-H because the cut-off points for SENV-H probe determined by our models seemed to be somewhat too low. We also obtained a low cut-off point for the SENV-C/H probe from the mixture model. This may be due to cross reactivity between genotypes of the assay with SENV-A or -D, caused by the choice of a relatively conserved region (Table I). These regions do not reflect the previous genetic distances on the phylogenetic analysis by Tanaka et al. [2001]. The low specificity of the SENV-C/H probe is also reflected in the high prevalence estimate in the US blood donor sample (7.8% for C/H versus 1.3% for SENV-A and 1.7% for SENV-D).

A general problem that arises when determining cut-off values for diagnostic tests is that cut-off points are population specific, and that the choice of diseased and comparative patient groups to estimate efficacy of a diagnostic test may bias sensitivity and specificity [Ransohoff and Feinstein, 1978]. One has to be cautious when applying cut-off points obtained from a highly infected population, such as IDUs, to a very low-risk population like the US blood donors. We thus validated the prevalence estimates that were obtained from applying the IDU sample cut-off points, by comparing them to prevalence estimates obtained from fitting the model directly to the blood donor data. While the cut-off points varied slightly, the prevalence estimates did not change.

The statistical methods, we used to obtain cut-off points for SENV infection can define positive cut points for any type of assay. Furthermore, if one has specific target sequences and a series of probes and each probe has a percent homology, then this information can be incorporated into the model to improve estimation of the optimal cut-off points for differentiating related genotypes. This approach can permit rapid and efficient detection of infection with a wide range of viruses.

APPENDIX

The Mixture Model

This statistical model is based on the premise that each person is in one of two latent true infection states, which we label as state $I = 1$ ("infected") and state $I = 0$ ("uninfected"). The probability density function of $y = \log_e (x + 0.5)$, where x denotes the chemiluminescence reading, is given by the mixture model (Model 1)

$$g(y|z, \theta) = f(y; \alpha_0)(1 - P) + f(y; \alpha_1)P, \qquad (1)$$

where $f(.;\alpha)$ is a density, such as the normal density, α denotes the parameters of f, and $\theta = (\alpha_0, \alpha_1, P)$. We interpret $f(.;\alpha_0)$ to be the density of the log transform of the chemiluminescence values that correspond to persons in the "uninfected" state, and $f(.;\alpha_1)$ to be the density of the values that corresponds to subjects in the "infected" state. The mixing probability P corresponds to the proportion of infected individuals in the population. The parameters in this model were estimated via a maximum likelihood procedure; for details see Pfeiffer et al. [1999]. An extension of this model let the state probabilities for the jth observation depend on covariates z_j, such as % homology, through a logistic regression:

$$P(I_j = 1|z_j) = P(z_j\beta) = \frac{\exp(z'_j\beta)}{(1 - \exp(z'_j\beta))}.$$

The first component of z_j was unity and corresponded to an intercept. β is a vector with k components of associated, unknown logistic regression coefficients. Given z, the probability density function of $y = \log(x + 0.5)$, where x denotes the chemiluminescence reading, is given by the mixture model, referred to as Model 2

$$g(y|z, \theta) = f(y;\alpha_0)(1 - P(z;\beta)) + f(y;\alpha_1)P(z;\beta). \qquad (2)$$

Calculation of the Cut-Off Points, Sensitivity, and Specificity

After Model (1) was fit to the data, and parameter estimates of θ were obtained, a cut-off value that "best" separates the two densities was calculated. Several optimality criteria were discussed in Pfeiffer et al. [1999]. To obtain the cut-off point that maximizes the sum of specificity (the probability that the test is negative given that the person is truly uninfected) $\int f(y;\alpha_0)dy$, and sensitivity $\int f(y; \alpha_1)\, dy$ (the probability that the test yield a positive result given that the person is truly infected), we found the value c^* that satisfied the equation $f(c^*;\alpha_0) - f(c^*; \alpha_1) = 0$.

Calculation of Posterior Probabilities of Infection Given % Homology

Model 2 allows one to calculate the probability of $I_j = 1$ ("infected") given the covariate z, the observed % homology, for each individual by calculating $P(I_j = 1|y, z) = f(y; \alpha_1)\, P(z; \beta)/\{f(y; \alpha_1)\, P(z; \beta) + f(y;\alpha_0)(1 - P(z; \beta))\}$ using the estimated parameter values θ, and the transformed chemiluminescence reading $y = \log(x + 0.5)$.

REFERENCES

Andreoletti L, Hober D, Belaich S, Lobert PE, Dewilde A, Wattre P. 1996. Rapid detection of enterovirus in clinical specimens using PCR and microwell capture hybridization assay. J Virol Methods 62:1–10.

Armitage P, Berry G. 1994. Statistical methods in medical research. 3rd edition. Oxford: Blackwell Scientific Publications.

Atkinson J, Edlin BR, Engels EA, Kral AH, Seal K, Gamache CJ, Whitby D, O'Brien TR. 2003. Seroprevalence of human herpesvirus 8 among infection drug users in San Francisco. J Infect Dis 187:974–981.

Hino S. 2002. TTV: A new human virus with single stranded circular DNA genome. Rev Med Virol 12(3):151–158.

Meden H, Marx D, Fattahi A, Rath W, Kron M, Wuttke W, Schauer A, Kuhn W. 1994. Elevated serum levels of a c-erbB-2 oncogene product in ovarian cancer patients and in pregnancy. J Cancer Res Clin Oncol 120(6):378–381.

Pfeiffer R, Gail MH, Brown L. 1999. A mixture model for the distribution of IgG antibodies to *Helicobacter pylori*: Application to studying factors that affect prevalence. J Epidemiol Biostat 5:267–275.

Primi DG, Fiordalisi GL, Mantero S, Mattioli A, Sottini F, Bonelli L, Vaglini M, Bonelli P, Olivero A. Dal Corso. 2000. Identification of SENV genotypes. International patent number WO 00/28039

(international application published under the patent cooperation treaty). Internet address: http://ep.espacenet.com/

Ransohoff DF, Feinstein AR. 1978. Problems of spectrum bias in evaluating the efficacy of diagnostic tests. N Engl J Med 17:926–929.

Rigas B, Hasan I, Rehman R, Donahue P, Wittkowski KM, Lebovics E. 2001. Effect on treatment outcome of coinfection with SEN viruses in patients with hepatitis C. Lancet 358:1961–1962.

Shibata M, Wang RY, Yoshiba M, Shih JW, Alter HJ, Mitamura K. 2001. The presence of a newly identified infectious agent (SEN virus) in patients with liver diseases and in blood donors in Japan. J Infect Dis 184:400–404.

Sottini AS, Mattioli G, Fiordalisi L, Mantero L, Imberti D, Moratto D, Primi D. 2001. In: Margolis H, editor. Molecular and biological characterization of SEN viruses: A family of viruses remotely related to the original TTV isolates. Proceedings of the 10th International Symposium on Viral Hepatitis and Liver Disease. International Medical Press.

Tanaka Y, Primi D, Wang RY, Umemura T, Yeo AET, Mizokami M, Alter HJ, Shih JWK. 2001. Genomic and molecular evolutionary analysis of a newly identified infectious agent (SEN virus) and its relationship to the TT virus family. J Infect Dis 183:359–367.

The Mathworks, Inc. 1999. Statistics toolbox user's guide.

Umemura T, Yeo AET, Sottini A, Moratto D, Tanaka Y, Wang RHY, Shih JYK, Donahue P, Primi D, Alter HJ. 2001. SEN virus infection and its relationship to transfusion-associated hepatitis. Hepatology 33:1303–1311.

Wilson LE, Umemura T, Astemborski J, Ray SC, Alter HJ, Strathdee SA, Vlahov D, Thomas DL. 2001. Dynamics of SEN virus infection among injection drug users. J Infect Dis 184:1315–1319.

Zhang P, Gebhart CJ, Burden D, Duhamel GE. 2000. Improved diagnosis of porcine proliferative enteropathy caused by Lawsonia intracellularis using polymerase chain reaction-enzyme-linked oligosorbent assay (PCR-ELOSA). Mol Cell Probes 14(2):101–108.